Nature Language Processing Assignment 3

24133101G WU Yifan

To design a QA System, I would like to define four main components.

Question Processing component: Analysis the question mentioned by user, extract keywords used define the type of questions, and the keywords for retrieval.

Input: The Questions Outputs: Types, Keywords

Document Retrieval component: Retrieval the set of documents based on keywords to find the most related documents and extract the most relevant passages from the set of documents.

Input: Keywords Outputs: (DocumentId, Passages)

Answer Extraction component: Extract the answer from the relevant passages directly.

Inputs: Type, (DocumentId, Passages) Output: Answer

Answer Generation component: For questions needed to inference, using generation model to generate answer with the related passages and types of questions.

Inputs: Type, (DocumentId, Passages) Output: Answer

An example of simple question of climate change:

Question: What is the immediate mechanism that causes glaciers to melt?

Document: The mechanism of glacier melting has been unclear in the past, but new research suggests that the 'ice albedo feedback' mechanism directly causes glacier melting. What is ice albedo feedback? When a glacier melts, the surface changes colour from white (snow and ice) to dark (rock or ocean), absorbing more solar radiation and accelerating warming in a vicious cycle.

Answer: ice albedo feedback

An example of complex question of climate change which needed to inference.

Question: What is the direct mechanism by which global warming causes glaciers to melt?

Document: In the past, it was widely believed that glacier melting was caused only by rising temperatures, but new research has shown that rising temperatures trigger an 'ice albedo feedback' mechanism: as glaciers melt, the surface changes colour from white (snow and ice) to darker (rock or ocean), absorbing more solar radiation and accelerating warming even further in a vicious cycle.

Answer: ice albedo feedback

The methods to process complex question:

Multi-hop Reasoning: Integrate answers from multiple passages through multiple rounds of searching and logical connections.

Retrieval-Augmented Generation: Use the search results as a context for generating models to generate comprehensive answers

Semantic Parsing Methods: Entity Linking -> Relation Detection -> Generate Structured Queries for answer generation. For example, SPARQL Query.